# Technical Consultant test case part 3

**Fransiskus Adrian Gunawan**

**29 September, 2022**

The following questions are designed to evaluate your thought process as you underwent the test. You may take your time to answer them, please be as descriptive as possible.

1. How long did you take to complete the test?

The assignment took me 2 hours and 15 minutes.

1. Was it easy or challenging? Which parts of the test were easy and which were challenging?

Initial expectations:

Initially, I find the test to be challenging. This was mainly due to my unfamiliarity with the domain and problem. In particular, I had to search external resources to understand the feature and target variables. I expect the translation of into code and building the ML model itself to be straightforward.

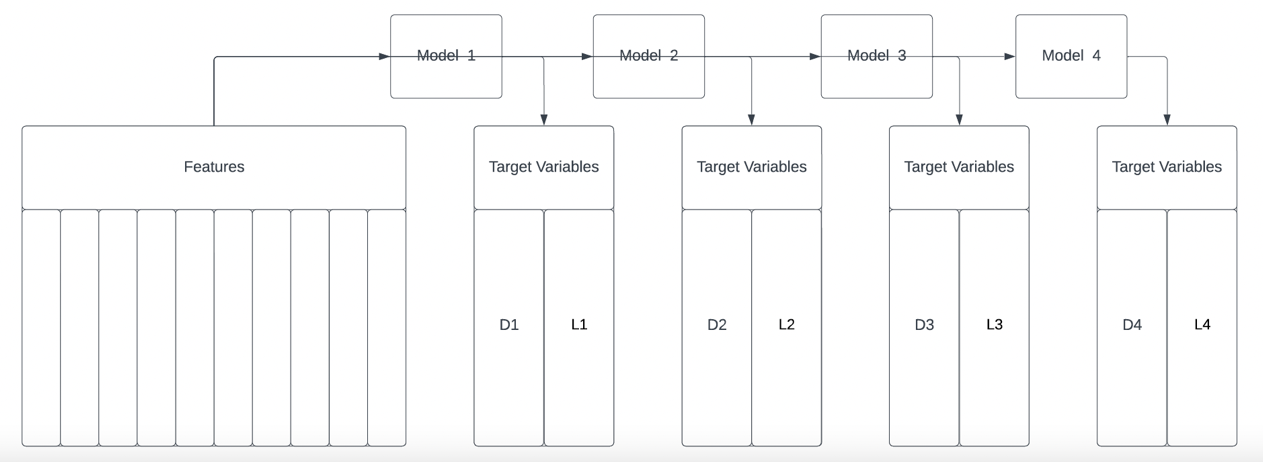
Impression post-test:

After completing the test, I realized that the most challenging part was understanding the clients requirements and performing feature selection/engineering. This was the case as I had to perform exploratory data analysis (i.e., data skewness, correlation with target, multicollinearity, number of unique values per feature). Although I could not finish this within the given timeframe, I expect the building of the model itself to be straightforward as I have worked on similar solutions previously.

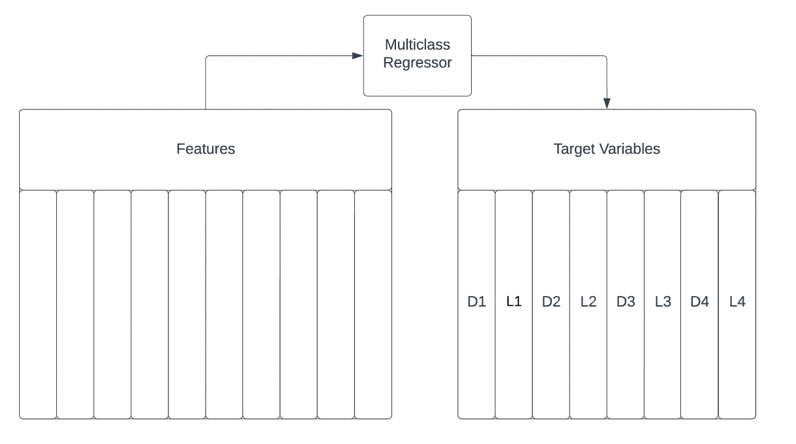
1. What resources did you use to learn how to solve the test (i.e. Google, forums, books)?

I mainly used Google to research the documentation of some packages (i.e., Sklearn) and to research the variables in the dataset. This was required as I was not very familiar with the terms/naming of each column, and the description provided was not sufficient to perform a thorough analysis of the data.

1. Briefly describe the process that you went through to find the solution for the problem.
   1. Initial heuristic approach (due to the limited time mindset) was to split data from each vessel separately and applying a separate ML model to each dataset. This can be visualized in the diagram below:



* 1. Upon further understanding of the data, the solution to the problem was straight-forward i.e., use one multioutput regressor on the entire dataset and predict the values for each of the ‘discharge’ and ‘load’ columns. This can be visualized in the diagram below:



Instead of building eight regressor models (one for each load and discharge), the multiclass regressor fits each target with one regressor; thus, only one model is required.

To test this approach, however, I would require a more detailed understanding of the data and each feature.

* 1. My typical approach for finding ML solutions regardless of the problem type is
     1. Get a basic understanding of the problem
     2. Find what type of ‘ML model’ could be used for the problem
     3. Build a basic pipeline with a simple ML model (relevant to the problem type)
     4. Get a deep understanding of the problem
     5. Do feature engineering
     6. Build the model
     7. Fine tune the model
     8. Repeat step v and vi until achieving an acceptable solution

1. Are there any other key experiences/notes that you would like us to know in regards to your experience in taking the test?

While I am familiar with the general workflow of similar ML projects, this was my first experience working with a multivariable regression problem. As such, there were several hurdles I faced, and approaches I had taken which might not be ideal for this problem. However, this test shows me that the knowledge I acquired regarding the data pre-processing and model building are still very much applicable in a new domain and problem type such as this.